SECTION 1-A3
PROTOCOL APPENDIX 3

STATISTICAL POWER FOR WOMEN'S HEALTH INITIATIVE CLINICAL TRIAL AND OBSERVATIONAL STUDY

1-A3.1 Clinical Trial (CT)

Introduction

As described in the main text there are three components of the CT: (A) dietary intervention, (B) hormone replacement therapy (HRT) and (C) calcium/vitamin D supplementation (CaD). Women who are eligible and willing to participate in either the dietary or the hormone component may enter the trial but each woman will be encouraged to participate in all components for which she is eligible. Thus a woman may participate in component A alone, B alone, A and B, A and C, B and C, or in all three components. We envisage that approximately 40% of the women who are willing and eligible to participate in the hormone replacement component will also be willing and eligible to participate in the dietary modification component. We also envisage that about 70% of women who enter the trial will accept participation in the CaD component.

In the following three subsections we describe the statistical power calculations for each of the three CT components. In the final subsection we pull together the results for each component and give a good sample size for the three components and the complete trial.

A. Dietary Intervention

The two main hypotheses of the dietary intervention component are that breast cancer incidence and colorectal cancer incidence will be reduced. A secondary hypothesis is that the incidence of coronary heart disease (CHD) will be reduced. Power calculations are presented for each of these hypotheses. Assumptions for these power calculations are described below.

Randomization Ratio. We propose to randomize 40% of the eligible women to dietary intervention and the remaining 60% to standard dietary advice. This ratio is chosen to reduce the costs of dietary intervention in the trial while maintaining statistical power.

Significance Level. Power is calculated for a two-sided test at significance level $\alpha = 0.05$. Note, however, that the preponderance of preliminary evidence argues against any increase in incidence of breast or colorectal cancer or CHD as a result of the dietary intervention.

Statistical Test. For breast cancer and colorectal cancer, we assume that a comparison of rates of new cases of disease in the two randomized groups, based upon a weighted logrank test, will be conducted as the main analysis. For CHD, we use a weighted binomial test, dividing the follow-up interval into three periods: 0-3 years, 3-6 years, and 6 or more years. This test accounts for the fact that complete data on CHD events will be available for individuals at three-yearly intervals, corresponding to the timing of the ECG examinations. The power calculations are performed using a modified version of a program written by Lakatos (Lakatos, *Biometrics*, 44, 229-241, 1988). For breast and colorectal cancer the weights increase linearly up to 10 years, corresponding to the model of Self *et al* (*Controlled Clinical Trials*, 9, 119-136, 1988). For CHD the weights for the three periods are one, two, and two, corresponding to an intervention effect that increases linearly up to three years and remains constant thereafter. These weights are chosen to maximize the statistical power of the study under time patterns of intervention effect that may be anticipated from dietary modification.

Trial Duration. Calculations are presented for average follow-up durations of six years and nine years.

Sample Size. Preliminary calculations indicated that adequate power would be obtained with a sample size of between 40,000 to 60,000. We present results for 42,000, 48,000 and 54,000 women.

Age Distribution. We assume women aged 50-54, 55-59, 60-69 and 70-79 years will enter the trial in the ratio 2:4:9:5.

Loss to Follow-up/Competing Risk. We assume a 3.0% per annum loss to follow-up due to deaths from other causes or disappearance, but only 2.0% per annum for the CHD endpoint, since about one third of deaths in the trial will be due to CHD.

Anticipated Intervention Effect.

(i) Breast Cancer. Self *et al (Controlled Clinical Trials,* 9, 119-136, 1988) discuss in detail the determination of the anticipated effect of a low-fat dietary intervention upon breast cancer incidence in a clinical trial. Their calculations involved assumptions regarding compliance, the magnitude of the dietary effect and the extent of the lag in this effect (described below). These assumptions led to an estimated 17% reduction in breast cancer incidence over an average follow-up period of approximately eight years.

Their compliance assumptions were based on data collected in the pilot study for the Women's Health Trial (1985-1986). The women in the intervention group in this trial decreased their average percent calories from fat from a baseline level of 39.5% to a level of 21.0% at six months, increasing to 21.7% at one year and to 22.6% at two years. It was assumed that their average level at 10 years would increase linearly to 26%. The women in the control group had corresponding average levels of 39.4%, 38.8%, 37.8% and 37.2% at baseline, six months, one year and two years. It was assumed that a total reduction of 4% in percent calories from fat would occur over the 10 years of the trial in these women. Early experience in the WHI indicates that the scenarios for the DM component of the CT had % energy from fat, as assessed by a food frequency instrument, that was considerably less than that reported in the original Women's Health Trial, which had taken place nearly a decade earlier. Hence women with a food frequency estimate of less than 32% energy from fat were excluded from the DM component, resulting in an estimated baseline average % energy from fat of about 35%. Daily fat gram goals for dietary intervention women were correspondingly reduced from those used in the preceding Women's Health Trial feasibility study in order to preserve an adequate difference in % energy from fat between control and intervention groups. This difference is assumed to be 13% at 1 year from randomization, dropping linearly to 11% over the subsequent nine years.

The magnitude of the dietary intervention effect was estimated from international correlations between dietary fat 'disappearance' data and breast cancer incidence rates. These data suggest a relative risk of 0.33 for lifetime consumption of 20% compared to 40% calories from fat. It was assumed that by the end of 10 years of intervention a relative risk of 0.50 would be achieved by fully compliant women, and that the relative risk would decrease linearly from 1.0 over this period. This gradually changing relative risk, from 1.0 at baseline to 0.5 at 10 years, leads, when averaged over the nine-year average follow-up period and taking non-compliance into account, to a projected 14% reduction in breast cancer incidence.

We calculate the power in the trial for detecting not only a 14% reduction but also a 12% or 11% reduction in breast cancer incidence corresponding to maximal 10 year relative risks of 44% and 39% respectively.

(ii) Colorectal Cancer. International correlations show similar relationships between total fat disappearance data and colorectal cancer incidence as with breast cancer incidence. The relationship between fat and colorectal cancer is supported by certain observational cohort studies (Willett, W. C., *et al, New England Journal of Medicine,* 323, 1664-1672, 1990). Moreover, some migrant studies suggest their more rapid adoption of indigenous incidence rates of colorectal cancer than of breast cancer (e.g. McMichael, A. J., Giles, G. G., *Cancer Research,* 48, 751-756, 1988). In addition, increases in dietary fiber, besides decreases in fat, are postulated to reduce colorectal cancer incidence. Thus, we may expect a greater effect of dietary intervention upon colorectal cancer incidence than upon breast cancer incidence, and a possibly shorter lag in this effect. For these reasons we anticipate a 9 year average reduction of approximately 18-22% in colorectal cancer incidence in this trial and present power calculations for 18%, 20% and 22% reductions.

(iii) Coronary Heart Disease. Data from the pilot study of the Women's Health Trial (Henderson *et al, Preventive Medicine*, 19**,** 115-133, 1990) indicate that the serum cholesterol levels of women in the intervention group fell on average by 6-7% over the initial two years. This occurred in the presence of the non-compliance levels described above. In addition results from the Lipid Research Clinics Coronary Primary Prevention Trial (LRC Program, *JAMA*, 251**,** 351-364, 1984) suggest that CHD incidence is reduced over a period of seven to eight years by approximately two times the percent reduction in serum cholesterol. Although the period of our trial is longer (nine years average follow-up), and we might therefore expect larger effects, we conservatively apply this two-fold factor to the 5-7% reduction in serum cholesterol achieved in the Women's Health Trial pilot study. Hence we investigate the power in this study to detect 9 year average reductions of 11%, 12% or 14% in CHD incidence. These values correspond to maximum full compliance effects of 21%, 24% and 26%.

Incidence and Mortality Rates.

(i) & (ii) Breast Cancer and Colorectal Cancer. We have used the age-specific incidence rates published from the SEER program for the years 1985-1989. No healthy volunteer correction is applied to the cancer incidence figures, since neither disease is thought to be strongly related to socioeconomic status or previous ill-health. Although there has been an increasing incidence of breast cancer in recent years, we conservatively assume that population incidence rates will remain steady during the trial.

(iii) Coronary Heart Disease. To account for the strong secular decrease in CHD-ID mortality (and, by inference, CHD morbidity) the published USA age-specific mortality rates for the years 1980-88 were projected forwards into the trial period, assuming that the linear trends would continue for a further 10 years and would thereafter stabilize. The incidence rates were obtained by multiplying the mortality rates by 2.5. They were then reduced by 33% to account for the healthy volunteer effect, i.e. the anticipation that women volunteering for the trial will have considerably lower CHD rates than in the general population.

(iv) Total Mortality. To illustrate the power of this study to ascertain the effect on total mortality we present a single table showing the power in the Dietary Modification (DM) component. Mortality rates were calculated for the control group by summing cause-specific mortality rates for breast cancer, colorectal cancer, CHD and death from other causes (see Loss to Follow-Up for rates). Cause-specific mortality rates were calculated from incidence rates in the following manner. For breast and colorectal cancer, mortality rates were estimated to be 25% and 35% of incidence rates respectively, based on SEER program and vital statistics information. For CHD mortality, the incidence rates were originally obtained from mortality rates so these calculations were inverted to obtain mortality rates adjusted for healthy volunteer effects and secular trends. Intervention effects were assumed to be the same for cause-specific mortality rates as for the associated incidence rates. Specifically 14%, 20% and 14% intervention affects were assumed for breast cancer, colorectal cancer and coronary heart disease, respectively. The table shows power for these design alternatives with varying degrees of effects on death from other causes (DOC). Note the dominating effect of deaths from other causes on the power of the total mortality comparison.

Statistical Power.

(i) Dietary Modification: Breast Cancer

Average follow-up (y)	Intervention effect (%)	Percentage of cases		Power (%) Sample size		
		Control	Intervention	42,000	48,000	54,000
6	11	2.05	1.89	25	28	31
	12	2.05	1.87	32	35	39
	14	2.05	1.85	39	44	48
9	11	2.92	2.61	58	63	69
	12	2.92	2.57	70	75	80
	14	2.92	2.52	81	86	89

☐ Power for design assumption highlighted.

(ii) Dietary Modification: Colorectal Cancer

Average follow-up (y)	Intervention effect (%)	Percentage of cases		Power (%) Sample size		
		Control	Intervention	42,000	48,000	54,000
6	18	1.07	0.93	33	37	41
	20	1.07	0.92	40	45	49
	22	1.07	0.91	47	52	57
9	18	1.61	1.32	77	83	87
	20	1.61	1.29	86	90	93
	22	1.60	1.25	92	95	97

(iii) Dietary Modification: CHD

Average follow-up (y)	Intervention effect	Percentage of cases		Power (%) Sample size		
		Control	Intervention	42,000	48,000	54,000
6	11	3.04	2.72	48	53	58
	12	3.03	2.67	58	64	69
	14	3.02	2.63	68	74	79
9	11	4.62	4.15	62	67	72
	12	4.60	4.08	72	78	83
	14	4.58	4.00	82	86	90

(iv) Dietary Modification: Total Mortality

Average follow-up (y)	Reduction in DOC	Percentage of cases		Power (%) Sample size
		Control	Intervention	48,000
6	0	7.59	7.33	18
	5%	7.59	7.06	58
9	0	12.85	12.41	29
	5%	12.85	11.92	85

B. Hormone Replacement Therapy

Women with a uterus will be randomized to one of two groups: Progestin/Estrogen Replacement Therapy (PERT) or Control (C). Those without a uterus will be randomized to Estrogen Replacement Therapy (ERT) or Control.

The main hypotheses are that ERT (or PERT) reduces CHD incidence. A secondary hypothesis is that ERT (or PERT) reduces the incidence of fractures. With regard to undesirable effects there is interest in testing for a possible increase due to ERT (or PERT) in the incidence of breast cancer. Estimated powers for each of these hypotheses are presented. Assumptions for these power calculations are described below.

The estimated power for comparing HRT to control, stratified by hysterectomy status, are presented for the same outcomes. These tests address the questions of the HRT strategy implemented as indicated by the woman's hysterectomy status.

Randomization Ratio. We propose to randomize 50% to PERT and 50% to Control among women with intact uteri; and 50% of the women without a uterus to ERT and 50% to Control. These ratios are chosen to maximize statistical power and reduce costs associated with monitoring women receiving ERT. We will restrict entry to ensure that approximately 45% of the women in the trial are post hysterectomy at randomization. The target hysterectomy rate is chosen to balance power for ERT vs. control and PERT vs. control while avoiding undue recruitment burden..

Significance Level. Power is calculated for a two-sided test at significance level $\alpha = 0.05$.

Statistical Test. As for dietary intervention for CHD. The weights for the weighted logrank tests for the effect on fractures are assumed to increase linearly up to three years of follow-up and remain constant thereafter. Separate tests will be made comparing ERT with Control and PERT with Control. The test for PERT will be interpreted in the context of the ERT test, since it is currently thought unlikely that PERT would achieve a

greater reduction in CHD incidence than ERT. Tests of any HRT vs. control will also be carried out for both CHD and fractures.

Trial Duration. As for dietary intervention.

Sample Size. Preliminary calculations indicated that adequate power for the main hypothesis would be obtained with a total sample size of between 20,000 to 30,000. We present results for 25,000, 27,500 and 30,000 women.

Age Distribution. As for dietary intervention.

Loss to Follow-up/Competing Risk. As for dietary intervention.

Drop-Out. We assume that in year one of follow-up 6% of women in the ERT (PERT) group will switch to Control and 3% per year thereafter up to year 10.

Drop-In. We assume that up to and including year five, 1.5% of women in the Control group per year will switch to ERT or PERT. For years 6 to 10 this annual rate will fall to 1.0%.

Lag. We assume that there will be a three-year lag in the effects of ERT or PERT on each of the disease incidences considered, except for breast cancer where the lag will be 10 years. Our assumption is that the treatment effect (on the hazard scale) increases linearly from zero at randomization to its full effect at the end of the lag period.

Anticipated Intervention Effect.

(i) & (ii) Coronary Heart Disease. As documented in the protocol, results from observational studies lead us to expect a reduction in incidence of approximately 50% in CHD for women taking ERT over a 10-year period. We assume that ERT and PERT will have similar intervention effects. Because of the likely existence of biases in the observational studies, we present estimated power for detecting a 25%, 30% or 35% reduction under full compliance, corresponding to an average 9 year reduction of 17%, 21% and 24% under adherence and lag time assumptions.

(iii) Fractures. Observational studies suggest that ERT may reduce hip fracture rates and "combined-site" fracture rates by 50%. Because of the likely existence of biases in observational studies, we present estimated power for detecting a 25%, 30%, or 35% reduction or an average reduction at 9 years of 18%, 21% and 25% under the above compliance and lag assumptions.

"Combined-site" fractures include those sites thought likely to fracture mainly due to osteoporosis in women over age 50 years. These are proximal femur (hip), distal forearm, proximal humerus, pelvis and vertebra. Hip is chosen as a separate endpoint because of the serious morbidity caused by this fracture.

(iv) Breast Cancer. Observational studies suggest that there may be an increased relative risk of breast cancer for women taking ERT. We present the estimated power for detecting a 20% or 30% increase under full compliance or a 14 year average increase of 15% and 22% after taking into account the projected noncompliance and lag time to effect.

Incidence Rates.

(i) Coronary Heart Disease. As for dietary intervention.

(ii) Fractures. Incidence figures of first fractures in five anatomical sites (proximal femur, distal forearm, proximal humerus, pelvis, vertebra) from Rochester, Minnesota study were provided by Dr. J. Melton. To obtain incidence for the combined sites, we summed the incidence figures for each individual site but then applied a multiplicative factor 0.8 to account for possible fractures at multiple sites that may occur in some women. We also applied a healthy volunteer effect of 0.8 to take account of the possible selection into the trial of women less likely to fracture. For hip fracture incidence we applied the same healthy volunteer correction.

(iii) Breast Cancer. As for dietary intervention. (Unlike dietary intervention, there may possibly be a healthy volunteer effect for breast cancer in the hormone replacement component due to some women with a family history of breast cancer being unwilling to participate. However, there is also an increasing rate of breast cancer in the general population. Neither of these phenomena are included in our calculations, but their effects are likely to cancel each other.)

***Statistical Power* (%).**

(i) ERT versus Control: CHD (assuming a hysterectomy rate of 45%)

Average Follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT	25,000	27,500	30,000
6	17	3.26	2.71	42	46	49
	21	3.26	2.60	57	62	65
	24	3.25	2.49	72	76	79
9	17	5.03	4.16	60	64	68
	21	5.02	3.97	77	81	84
	24	5.01	3.79	89	92	94

(ia) PERT versus Control: CHD (assuming intact uterus in 55%)

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	PERT	25,000	27,500	30,000
6	17	3.26	2.71	50	54	57
	21	3.26	2.60	66	70	74
	24	3.25	2.49	80	84	88
9	17	5.03	4.16	69	73	77
	21	5.02	3.97	85	88	90
	24	5.01	3.79	94	96	97

(ib) Combined ERT/PERT: CHD

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT/PERT	25,000	27,500	30,000
6	17	3.26	2.71	75	79	82
	21	3.26	2.60	89	92	94
	24	3.25	2.49	97	98	99
9	17	5.03	4.16	91	94	95
	21	5.02	3.97	98	99	99
	24	5.01	3.79	>99	>99	>99

(ii) ERT versus Control: Combined Fractures (assuming a hysterectomy rate of 45%)

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT	25,000	27,500	30,000
6	17	7.83	6.55	80	84	87
	20	7.82	6.29	93	95	96
	23	7.81	6.02	98	99	99
9	17	11.85	9.88	94	96	97
	20	11.83	9.46	99	99	>99
	23	11.80	9.03	>99	>99	>99

(iia) PERT versus Control: Combined Fractures (assuming intact uterus in 55%)

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	PERT	25,000	27,500	30,000
6	17	7.83	6.55	88	90	93
	20	7.82	6.29	97	98	99
	23	7.81	6.02	99	>99	>99
9	17	11.85	9.88	97	98	99
	20	11.83	9.46	>99	>99	>99
	23	11.80	9.03	>99	>99	>99

(iib) Combined ERT/PERT: Combined Fractures

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT/PERT	25,000	27,500	30,000
6	25	7.83	6.55	99	99	>99
	30	7.82	6.29	>99	>99	>99
	35	7.81	6.02	>99	>99	>99
9	25	11.85	9.88	>99	>99	>99
	30	11.83	9.46	>99	>99	>99
	35	11.80	9.03	>99	>99	>99

(iii) ERT versus Control: Hip Fractures (assuming a hysterectomy rate of 45%)

Average Follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT	25,000	27,500	30,000
6	18	1.88	1.55	29	31	34
	21	1.87	1.49	40	43	46
	25	1.87	1.42	52	56	60
9	18	3.14	2.58	45	48	52
	21	3.13	2.46	60	65	68
	25	3.12	2.35	75	79	82

(iiia) PERT versus Control: Hip Fractures (assuming intact uterus in 55%)

Average Follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	PERT	25,000	27,500	30,000
6	18	1.88	1.55	34	37	40
	21	1.87	1.49	47	51	54
	25	1.87	1.42	60	65	68
9	18	3.14	2.58	52	56	60
	21	3.13	2.46	69	73	77
	25	3.12	2.35	83	86	89

(iiib) Combined ERT/PERT: Hip Fracture

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT/PERT	25,000	27,500	30,000
6	18	1.88	1.55	55	59	63
	21	1.87	1.49	72	76	79
	25	1.87	1.42	85	88	91
9	18	3.14	2.58	78	82	85
	21	3.13	2.46	91	94	95
	25	3.12	2.35	98	98	99

(iv) ERT versus Control: Breast Cancer (assuming a hysterectomy rate of 45%)

Average Follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT	25,000	27,500	30,000
6	15	2.07	2.25	11	12	12
	22	2.08	2.35	19	21	22
9	15	3.04	3.40	22	24	26
	22	3.05	3.60	43	46	50
14	15	4.53	5.21	44	47	50
	22	4.56	5.58	75	79	83

(iva) PERT versus Control: Breast Cancer (assuming intact uterus in 55%)

Average Follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	PERT	25,000	27,500	30,000
	15	2.07	2.25	13	13	14
6	22	2.08	2.35	23	24	26
	15	3.04	3.40	26	29	31
9	22	3.05	3.60	59	54	58
	15	4.53	5.21	51	55	59
14	22	4.56	5.58	83	87	89

(ivb) Combined ERT/PERT: Breast Cancer

Average follow-up (y)	Intervention effect (%)	Percentage cases		Power (%) Total sample size		
		Control	ERT/PERT	25,000	27,500	30,000
6	15	2.07	2.25	19	21	22
	22	2.08	2.35	37	40	43
9	15	3.04	3.40	43	47	50
	22	3.05	3.60	76	79	83
14	15	4.53	5.21	76	80	84
	22	4.56	5.58	98	99	99

C. Calcium/Vitamin D

Calcium/Vitamin D. Women will be randomized to one of two groups: CaD supplementation or Placebo. The main hypothesis is that CaD reduces fracture rates.

Estimated powers for the main hypothesis are greatly in excess of those calculated for the effect of ERT on fracture rates, as shown below.

Randomization Ratio. We will randomize 50% of eligible women to CaD and 50% to Placebo.

Significance Level. As for hormone replacement.

Statistical Test. As for hormone replacement.

Trial Duration. Average follow-up for this component will be one year less than for the other components. We tabulate powers for five years and eight years follow-up.

Sample Size. We calculate power for sample sizes of 25,000, 35,000, and 45,000.

Age Distribution. As above.

Loss to Follow-Up/Competing Risk. As above.

Drop-Out and Drop-In. These rates should be considerably lower than for hormone replacement. However we assume, conservatively, that they are the same.

Lag. As for Hormone Replacement for Fractures, and 10 years for Colorectal Cancer.

Anticipated Effect. Observational studies suggest the same level of effect on the incidence of fractures as ERT. We therefore adopt the same effects as for hormone replacement. We calculate power for a reduction of 18-22% in the incidence of colorectal cancer.

Incidence Rates. As above.

Statistical Power.

(i) CaD versus Control: Hip Fractures

		Percentage cases		Power (%) Total sample size		
Average Follow-up (y)	Intervention effect (%)	Control	Calcium/ Vitamin D	25,000	35,000	45,000
5	18	1.51	1.26	44	57	68
	21	1.51	1.21	60	74	84
	25	1.51	1.16	74	87	94
8	18	2.69	2.21	72	85	93
	21	2.68	2.11	87	95	99
	25	2.68	2.01	95	99	>99

(ii) CaD versus Control: Combined Fractures

		Percentage cases		Power (%) Total sample size		
Average Follow-up (y)	Intervention effect (%)	Control	Calcium/ Vitamin D	25,000	35,000	45,000
5	17	6.50	5.49	96	99	>99
	20	6.50	5.28	99	>99	>99
	24	6.49	5.07	>99	>99	>99
8	17	10.50	8.75	>99	>99	>99
	20	10.48	8.38	>99	>99	>99
	24	10.46	8.00	>99	>99	>99

(iii) CaD versus Control: Colorectal Cancer

		Percentage cases		Power (%) Total sample size		
Average Follow-up (y)	Intervention effect (%)	Control	Calcium/ Vitamin D	25,000	35,000	45,000
5	18	0.86	0.76	17	23	28
	20	0.86	0.75	20	27	33
	22	0.86	0.74	24	31	38
8	18	1.42	1.17	52	66	77
	20	1.42	1.15	60	75	85
	22	1.42	1.12	69	83	91

Conclusion

From these power calculations we propose a trial with an average nine years follow-up and the following sample size:

(A) For the DM component, 48,000 women

(B) For the HRT component, 27,500 women

(C) For the CaD component, 45,000 women.

These sample sizes provide substantial power for all the main hypotheses: Diet versus Breast Cancer (86% for a 14% average reduction), Diet versus Colorectal Cancer (90% for a 20% average reduction), HRT versus CHD (81-88% for a 21% average reduction), CaD versus Fractures (93% for a 21% average reduction in Hip Fractures) and substantial power also for some secondary hypotheses: Diet versus CHD (86% for a 14% average reduction), Hormone Replacement versus Fractures (65-73% for a 21% average reduction in Hip Fractures) and Calcium versus Colorectal Cancer (85% for a 20% average reduction).

There is 46-54% power for detecting a 22% average increase in breast cancer risk (or a 30% increase with full compliance) from either HRT preparation should it exist, and a 29% possibility of detecting even a 14% (20% full compliance) increase. A further five years follow-up of women in the hormone replacement component is planned to increase the power in monitoring for this potential adverse effect. The power for detecting a 22% increase after 14 years average follow-up is 79-87%, and for a 14% increase is 47-55%.

According to our estimate that 40% of the 27,500 women participating in the hormone component will also participate in the dietary component, the total number of women anticipated in the trial will be 64,500.

1-A3.2 Observational Study

There are a number of factors to be considered in describing the power of the OS to elucidate relationships between baseline measurements and subsequent disease risk, as well as relationships between changes in measurements from baseline and three years and subsequent disease risk. These include:

(i) Incidence rates for diseases of interest - as described in *Section 1* of the Protocol and in the earlier part of this Appendix, incidence rates are quite variable for the diseases of interest in the Women's Health Initiative (WHI). For example, the annual incidence rates for some key outcome categories, assuming that 10%, 20%, 45% and 25% of OS enrollees are in the age categories 50-54, 55-59, 60-69 and 70-79, respectively, are approximately 5.0 for CHD, 3.0 for breast cancer, 1.8 for colorectal cancer, and 4.0 for hip fractures, per 1,000 enrollees. Naturally, it will be desirable to use the OS for studies of less common outcomes, including specific cancers (e.g., endometrial, ovarian), selected vascular diseases (e.g., hemorrhagic stroke, deep vein thrombosis), and fractures at specific, less common sites. The annual incidence rates for such diseases may be

less than 1.0, or even less than 0.5 per thousand. Hence, generic power calculations have been conducted for annual incidence rates of 0.1, 0.5, 1.0, 2.0 and 5.0 per thousand.

(ii) Follow-up durations - it is particularly important that the OS begin to generate research reports as early as possible during the course of the WHI program. Hence, power calculations have been performed for average cohort follow-up durations of 3, 6 and 9 years. The three-year power calculations, for example, can be applied to studies of baseline characteristics when the average follow-up time for the OS (or a subset thereof) is three years, or to the study of changes in characteristics between baseline and three years when the average follow-up time is six years, since outcomes prior to a participants three-year visit do not contribute to these latter analyses.

(iii) Sample size and subset analyses - power calculations based on the entire intended OS sample size of 100,000 are perhaps of most interest, but there is also considerable interest in analyses based on various OS subsets. For example, separate analyses for each decade of baseline age would require power calculations for cohorts in the range of 25,000 to 45,000 subjects in view of the anticipated OS age distribution mentioned above. Similarly, the anticipated OS enrollment by racial/ethnic subgroup is as follows: non-Hispanic, white - 80,000; African American - 10,000; Hispanic - 6,000; Native American - 2,000; Asian Pacific Islander - 2,000. Other analyses may be restricted to OS women for whom a certain measurement falls within selected percentiles relative to the overall OS distribution. For example, an important goal of the OS pertains to further elucidation of the relationship between a low-blood cholesterol or a recent reduction in blood cholesterol and subsequent mortality. Analyses restricted to the approximately 40,000 women with baseline blood cholesterol in the lowest two quintiles may provide particular insights. For example, one will be able to compare the mortality rates of women with blood cholesterol measurements in the lowest quintile at both baseline and three years, to those whose cholesterol has dropped from the second lowest to the lowest quintile between baseline and their three-year visit.

Power calculations were conducted for sample sizes of 100,000; 80,000; 40,000; 20,000; 10,000; 6,000 and 2,000 in order to explore the relationship between power and subset sample size.

(iv) Distribution of exposures or characteristics - the characteristics or exposures to be related to disease risk may involve a variety of types of measurements, including binary, categorical and continuous variates, and mixtures thereof. However, most analyses, especially exploratory analyses, will involve the comparison of disease risks between two groups of OS members distinguished by their values of one or more characteristics. For example, one may compare current ERT users to non-users; or may compare women in the highest quintile of baseline blood cholesterol, or baseline dietary fat intake, to corresponding women in the lowest quintile. Hence, power calculations were conducted as a function of the frequency of a binary characteristic or exposure, with 'exposure' frequencies taking values of 0.5%, 1%, 10%, 30% and 50%. For example, to obtain the power of a comparison of the highest quintile to lowest quintile of blood cholesterol in the entire OS cohort one can examine the following tables for a sample size of 40,000 (the highest and lowest quintiles combined) with an exposure frequency of 50% (one-half of the 40,000 women will be in the highest quintile).

(v) Odds ratio - there are a range of odds ratio values that may be pertinent to associations of interest in the OS. Odds ratios of 2.0 or above may have particular public health importance, particularly if the characteristic under study is fairly common. Note that odds ratios and relative risks are virtually identical for the range of disease incidence rates mentioned above.

In considering the range of odds ratios pertinent to the OS, it is important to consider the regression attenuation that arises from random measurement error in the assessment of characteristics of interest. For example, the slope of the regression line that relates the log-disease incidence (e.g., log-CHD incidence) to a single blood cholesterol measurement are attenuated by a factor of about 2/3 on the basis of such random measurement error, so that an odds ratio of 2 is reduced to $\exp\{(2/3)\log 2\}=1.59$ by (non-differential) measurement error. The corresponding attenuation factor for estimates of nutrient intakes based on a food frequency instrument may be in the vicinity of 1/3 depending upon the nutrient and assessment instrument, so that an odds ratio of 2 is attenuated to about 1.26 based on random measurement error for such exposures. Hence, to explore the power of the OS under various configurations of association strength and regression dilution, power calculations have been conducted for odds ratios of 1.25, 1.50, 1.75, 2.0 and 3.0.

(vi) Sampling procedures, and confounding factor control - the power calculations that follow assume the characteristic or exposure under study to be available on all pertinent study subjects, and uses the asymptotic distribution of a simple odds ratio statistic. However, many of the OS analyses will use time-matched case-control, or stratified case-cohort, sampling to reduce the number of women for whom expensive analysis of stored specimens or complicated questionnaires must be carried out. The efficiency of a time-matched case-control analysis as compared to a full cohort analysis is approximately $k(k+1)^{-1}$, where k is the number of controls matched to each case. Hence, a one-to-k matched case-control study based on a cohort of size n has power approximately equal to a full-cohort analyses based on a sample of size $nk(k+1)^{-1}$.

The following array can be used to approximately convert full-cohort sample size to corresponding 1:k matched case-control effective sample size for k= *1,2,3,5*.

Effective Cohort Sizes for 1:k Matched Case-Control Analysis

Full Cohort Sample Sizes

Controls (k)

per case	100,000	80,000	40,000	20,000	10,000	6,000	2,000
1	50,000	40,000	20,000	10,000	5,000	3,000	1,000
2	66,667	53,333	26,667	13,333	6,667	4,000	1,333
3	75,000	60,000	30,000	15,000	7,500	4,500	1,500
4	80,000	64,000	32,000	16,000	8,000	4,800	1,600
5	83,333	66,667	33,333	16,667	8,333	5,000	1,667

Most OS analyses will also make provision, via stratification, matching or regression modeling, for factors that have potential to confound the association under study. Such control is essential to accurate odds ratio, or relative risk estimation, and corresponding more complex tests will tend to have reduced power, relative to the corresponding test in which confounding control is unnecessary. However, the power reduction is likely to be quite minor in most OS analyses so that no provision for confounding control is included in the OS power calculations.

The following tables present the power calculations for the configurations listed above, with the exception that combinations of factors for which the power is less than 50% are omitted for brevity.

Table 1-A3.1. provides power calculations for analyses based on the entire cohort of size 100,000. For example, from the lower section of *Table 1-A3.1.* one can see that the power for detecting a relative risk of 1.5 associated with a characteristic present in 50% of the cohort is 72% after an average three years of follow-up, and 95% after an average of six years of follow-up, even for a disease with annual disease incidence of .05% per year, which is close to that for cancers of the endometrium and ovary, for example, An odds ratio of about 1.5 for above versus below the median fat intake can be projected from international correlation analyses for endometrial and ovarian cancer, after accounting for regression dilution. Similarly, an odds ratio of 2.0 associated with a characteristic arising in only 1% of the cohort can be detected with adequate power for diseases as common as breast cancer or hip fractures, and can be detected with power 83% after an average of only three years of follow-up for a disease such as CHD having an annual incidence of about .5% per year or greater.

Table 1-A3.2. presents corresponding analyses for a subsample of the OS of size 80,000. As such, it gives projected power for OS analyses restricted to non-Hispanic white women or for analyses on the entire 100,000 women based on a case-control analysis with four controls per case. Note that the power reductions in moving

from *Table 1-A3.1.* to *Table 1-A3.2.* tend to be fairly modest. Consider two specific associations which could be examined in the OS: About 5-10% of postmenopausal women have serum ferritin concentrations about 200 µg/liter. A study in Finnish men indicates that such elevated concentrations may convey an odds ratio of about 2.2 for CHD. *Table 1-A3.2.* indicates that a 1:4 matched case-control study in the OS cohort would have power in the vicinity of 90% for detecting an elevated serum ferritin and CHD association, even if the odds ratio is as small as 1.25. As a second example, suppose that a particular occupational group, such as a lab technician or hair dresser, constitutes only .5% of the OS cohort. *Table 1-A3.2.* indicates that a 1:4 matched case-control study based on the OS would have power of at least 76% by an average six years of follow-up, or 94% by an average of nine years of follow-up, for detecting an odds ratio of 3.0 for a disease such as breast cancer with an annual incidence rate of two per 1,000 or greater. In fact, a British Columbia study suggests a breast cancer odds ratio of about four for these occupational groups.

Table 1-A3.3. shows corresponding power calculations for a subsample of size 40,000, as corresponds, for example, to studies restricted to extreme quintiles of a measured characteristic. A relative risk as small as 1.50 between extreme quintiles of a nutrient intake variable, for example, will be able to be detected with power 90% or greater by an average of three years of follow-up for diseases such as breast cancer, hip fractures or CHD having an annual incidence of at least .2%. Such an odds ratio can be detected with a power of 80% for a much rarer disease with incidence of .05% per year, by an average of nine years of follow-up. *Table 1-A3.4.* gives corresponding power calculations for a subsample of size 20,000. These entries are pertinent to full-cohort analyses restricted to the subset of women in the age range 70-79 at baseline, and to subsamples of size 40,000 under 1:1 matched case-control sampling.

Table 1-A3.5. gives power calculations for a subsample of size 10,000 - the anticipated number of African Americans in the OS. Note that there will be adequate power to detect an odds ratio of 1.50 or larger for diseases of annual incidence of .2% or larger, provided the characteristics or exposure arises in about half of the women in the subsample. *Table 1-A3.6.* gives power calculations for a subsample of size 6,000 - the anticipated number of Hispanic American women in the OS. There is adequate power to detect an odds ratio of 1.75 or larger for diseases of annual incidence of .2% per year or larger, again provided the characteristic arises in about 50% of the subsample. Finally, *Table 1-A3.7.* gives power calculations for a subsample of size 2,000 - the anticipated number of Native American, and of Asian and Pacific Islander American women in the OS. Odds ratios of 3.0 will be able to be detected for diseases having annual incidence of about .2% per year or greater, provided the characteristic under study arises in about 50% of the subsample.

Table 1-A3.1
OS Power Calculations for Cohort Size of 100,000

		Annual Disease Incidence Per 1,000 Women														
		0.1			**0.5**			**1.0**			**2.0**			**5.0**		
Average Years of Follow-up		**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**
Exposure Frequency	Odds Ratio															
0.50%	1.75															0.79
	2.00												0.55		0.82	0.95
	3.00									0.73		0.88	0.98	0.96	1.00	1.00
1.00%	1.50														0.60	0.79
	1.75												0.69	0.59	0.91	0.98
	2.00									0.55		0.72	0.90	0.83	0.99	1.00
	3.00						0.72		0.89	0.99	0.89	1.00	1.00	1.00	1.00	1.00
10%	1.25									0.52		0.65	0.83	0.75	0.96	1.00
	1.50					0.57	0.77	0.57	0.89	0.98	0.89	1.00	1.00	1.00	1.00	1.00
	1.75				0.55	0.89	0.98	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00			0.51	0.78	0.97	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00		0.83	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30%	1.25						0.59		0.72	0.88	0.72	0.95	0.99	0.98	1.00	1.00
	1.50				0.64	0.92	0.99	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.75		0.51	0.71	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00		0.73	0.90	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50%	1.25						0.66		0.8	0.93	0.80	0.98	1.00	0.99	1.00	1.00
	1.50				0.72	0.95	0.99	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.75		0.59	0.78	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00		0.80	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1-A3.2
OS Power Calculations for a Subsample Size of 80,000

		Annual Disease Incidence Per 1,000 Women														
		0.1			0.5			1.0			2.0			5.0		
Average Years of Follow-up		3	6	9	3	6	9	3	6	9	3	6	9	3	6	9
Exposure Frequency	Odds Ratio															
0.50%	1.75															0.67
	2.00														0.71	0.89
	3.00									0.57		0.76	0.94	0.88	1.00	1.00
1.00%	1.50															0.69
	1.75												0.57		0.83	0.95
	2.00											0.59	0.81	0.72	0.97	1.00
	3.00						0.57		0.77	0.95	0.77	0.99	1.00	1.00	1.00	1.00
10%	1.25											0.54	0.73	0.64	0.92	0.98
	1.50						0.66		0.80	0.94	0.80	0.98	1.00	1.00	1.00	1.00
	1.75					0.79	0.94	0.79	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00
	2.00				0.66	0.95	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00		0.69	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30%	1.25								0.62	0.80	0.62	0.90	0.98	0.95	1.00	1.00
	1.50				0.53	0.85	0.96	0.85	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	1.75			0.60	0.84	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00		0.62	0.82	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00	0.75	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50%	1.25						0.56		0.69	0.86	0.69	0.94	0.99	0.98	1.00	1.00
	1.50				0.61	0.90	0.98	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.75			0.68	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00		0.69	0.87	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1-A3.3
OS Power Calculations for a Subsample Size of 40,000

		Annual Disease Incidence Per 1,000 Women														
		0.1			0.5			1.0			2.0			5.0		
Average Years of Follow-up		3	6	9	3	6	9	3	6	9	3	6	9	3	6	9
Exposure Frequency	Odds Ratio															
0.50%	2.00															0.53
	3.00												0.57		0.87	0.98
1.00%	1.75															0.67
	2.00														0.70	0.89
	3.00									0.57		0.76	0.94	0.88	1.00	1.00
10%	1.25														0.64	0.81
	1.50									0.66		0.79	0.93	0.88	1.00	1.00
	1.75						0.64		0.79	0.94	0.79	0.98	1.00	1.00	1.00	1.00
	2.00					0.66	0.86	0.66	0.95	1.00	0.95	1.00	1.00	1.00	1.00	1.00
	3.00			0.51	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30%	1.25											0.61	0.79	0.71	0.95	0.99
	1.50					0.53	0.73	0.53	0.85	0.96	0.85	0.99	1.00	1.00	1.00	1.00
	1.75				0.51	0.84	0.96	0.84	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	2.00				0.73	0.97	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00		0.76	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50%	1.25									0.56		0.69	0.86	0.79	0.97	1.00
	1.50					0.61	0.80	0.61	0.90	0.98	0.90	1.00	1.00	1.00	1.00	1.00
	1.75				0.59	0.89	0.98	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.00			0.55	0.80	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3.00		0.81	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1-A3.4
OS Power Calculations for a Subsample Size of 20,000

		Annual Disease Incidence Per 1,000 Women														
		0.1			**0.5**			**1.0**			**2.0**			**5.0**		
Average Years of Follow-up		**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**	**3**	**6**	**9**
Exposure Frequency	Odds Ratio															
0.50%	3.00															0.69
1.00%	2.00															0.53
	3.00												0.56		0.86	0.98
10%	1.25															0.50
	1.50												0.65	0.56	0.87	0.97
	1.75									0.64		0.79	0.93	0.88	1.00	1.00
	2.00						0.50		0.66	0.86	0.66	0.95	1.00	0.98	1.00	1.00
	3.00					0.82	0.97	0.82	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
30%	1.25														0.71	0.87
	1.50								0.53	0.72	0.53	0.85	0.96	0.92	1.00	1.00
	1.75					0.51	0.81	0.51	0.84	0.96	0.84	0.99	1.00	1.00	1.00	1.00
	2.00					0.73	0.90	0.73	0.97	1.00	0.97	1.00	1.00	1.00	1.00	1.00
	3.00			0.59	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50%	1.25												0.56		0.78	0.92
	1.50								0.61	0.79	0.61	0.90	0.98	0.95	1.00	1.00
	1.75					0.59	0.78	0.59	0.89	0.98	0.89	1.00	1.00	1.00	1.00	1.00
	2.00					0.80	0.94	0.80	0.98	1.00	0.98	1.00	1.00	1.00	1.00	1.00
	3.00			0.67	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1-A3.5
OS Power Calculations for a Subsample Size of 10,000

		Annual Disease Incidence Per 1,000 Women											
		0.5			1.0			2.0			5.0		
Average Years of Follow-up		3	6	9	3	6	9	3	6	9	3	6	9
Exposure Frequency	Odds Ratio												
1.00%	3.00												0.69
10%	1.50												0.75
	1.75									0.63	0.54	0.87	0.97
	2.00						0.50		0.66	0.86	0.77	0.98	1.00
	3.00			0.65		0.82	0.96	0.82	0.99	1.00	1.00	1.00	1.00
30%	1.25												0.57
	1.50								0.53	0.72	0.63	0.91	0.98
	1.75					0.51	0.71	0.51	0.84	0.96	0.91	1.00	1.00
	2.00			0.58		0.73	0.90	0.73	0.97	1.00	0.99	1.00	1.00
	3.00		0.86	0.97	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50%	1.25												0.64
	1.50								0.61	0.79	0.71	0.95	0.99
	1.75					0.59	0.78	0.59	0.89	0.98	0.95	1.00	1.00
	2.00			0.66		0.80	0.93	0.80	0.98	1.00	1.00	1.00	1.00
	3.00	0.58	0.90	0.98	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1-A3.6.
OS Power Calculations for a Subsample Size of 6,000

		Annual Disease Incidence Per 1,000 Women											
		0.5			1.0			2.0			5.0		
Average Years of Follow-up		3	6	9	3	6	9	3	6	9	3	6	9
Exposure Frequency	Odds Ratio												
10%	1.50												0.50
	1.75											0.63	0.82
	2.00									0.60	0.50	0.85	0.96
	3.00					0.51	0.76	0.51	0.90	0.99	0.96	1.00	1.00
30%	1.50											0.71	0.88
	1.75								0.60	0.79	0.71	0.95	0.99
	2.00						0.67		0.81	0.95	0.90	1.00	1.00
	3.00		0.59	0.81	0.59	0.92	0.99	0.92	1.00	1.00	1.00	1.00	1.00
50%	1.50									0.56		0.78	0.92
	1.75						0.53		0.67	0.85	0.78	0.97	1.00
	2.00					0.55	0.75	0.55	0.87	0.97	0.93	1.00	1.00
	3.00		0.67	0.86	0.67	0.95	0.99	0.95	1.00	1.00	1.00	1.00	1.00

Table 1-A3.7
OS Power Calculations for a Subsample Size of 2,000

		Annual Disease Incidence Per 1,000 Women								
		1.0			2.0			5.0		
Average Years of Follow-up		3	6	9	3	6	9	3	6	9
Exposure Frequency	Odds Ratio									
10%	3.00						0.50		0.80	0.95
30%	1.75								0.50	0.69
	2.00								0.71	0.88
	3.00			0.59		0.75	0.92	0.85	0.99	1.00
50%	1.75								0.58	0.76
	2.00						0.55		0.78	0.92
	3.00			0.67		0.81	0.94	0.89	1.00	1.00

Section 1-A3
Protocol Appendix 3
Statistical Power for Women's Health Initiative
Clinical Trial and Observational Study

Table of Contents

Contents **Page**